

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (2/98)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No. 042390.P9663
First Inventor or Application Identifier Ganapati Srinivasa
Title RUNTIME PREDICTION FRAMEWORK FOR CPU INTENSIVE APPLICATIONS
Express Mail Label No. EL634500832US

APPLICATION ELEMENTS
See MPEP chapter 600 concerning utility patent application contents

ADDRESS TO: Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

1. ☒ Fee Transmittal Form (e.g. PTO/SB/17)
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification *Total Pages* 12
(preferred arrangement set forth below)
- Descriptive title of the Invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 U.S.C.113) *Total Sheets* 4
4. Oath or Declaration *Total Pages* ☐
- a. ☐ Newly executed (original copy)
- b. ☐ Copy from a prior application (37 CFR 1.63(d))
(for continuation/divisional with Box 16 completed)
- i. ☐ **DELETION OF INVENTOR(S)**
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).

5. ☐ Microfiche Computer Program (Appendix)
6. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
- a. ☐ Computer Readable Copy
- b. ☐ Paper Copy (identical to computer copy)
- c. ☐ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

7. ☐ Assignment Papers (cover sheet & document(s))
8. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney
(when there is an assignee)
9. ☐ English Translation Document (if applicable)
10. ☐ Information Disclosure Statement (IDS)/PTO - 1449 ☐ Copies of IDS Citations
11. ☐ Preliminary Amendment
12. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
13. ☐ *Small Entity ☐ Statement filed in prior application, Status still proper and desired
14. ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
15. ☐ Other:

*NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).

16. If a **CONTINUING APPLICATION**, check appropriate box, and supply the requisite information below and in a preliminary amendment:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: _____/_____

Prior application Information: Examiner _____ Group/Art Unit: _____

For **CONTINUATION** or **DIVISIONAL APPS** only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

17. CORRESPONDENCE ADDRESS

☐ Customer Number of Bar Code Label

(Insert Customer No. or Attach bare code label here)

or ☒ Correspondence address below

Name	BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP				
Address	12400 Wilshire Boulevard, Seventh Floor				
City	Los Angeles	State	California	Zip Code	90025
Country	U.S.A.	Telephone	(310) 207-3800	Fax	(310) 820-5988

Name (Print/Type) Farzad E. Amini, Reg. 42,261

Signature

Date

09/29/00

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

FEE TRANSMITTAL

Patent fees are subject to annual revision on October 1.
These are the fees effective October 1, 1997.
Small Entity payments must be supported by a small entity statement,
otherwise large entity fees must be paid. See Forms PTO/SB/09-12.
See 37 C.F.R. §§ 1.28 and 1.29

TOTAL AMOUNT OF PAYMENT (\$) **690.00**

Complete if Known

Application Number	
Filing Date	09/29/00
First Named Inventor	Ganapati Srinivasa, et al.
Examiner Name	
Group Art Unit	
Attorney Docket Number	042390.P9663

METHOD OF PAYMENT (check one)

1. ☒ The Commissioner is hereby authorized to charge indicated fees and credit any over payments to:

Deposit Account Number **02-2666**

Deposit Account Name **Blakely, Sokoloff, Taylor & Zafman LLP**

- ☒ Charge Any Additional Fee Required Under 37 CFR 1.16 and 1.17 ☐ Charge the Issue Fee Set in 37 CFR 1.18 at the Mailing of the Notice of Allowance.

2. ☒ Payment Enclosed:

- ☒ Check ☐ Money Order ☐ Other

FEE CALCULATION (fees effective 10/01/96)

1. FILING FEE

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description	Fee Paid
101 690	201 345	Utility filing fee	\$690
106 310	206 155	Design filing fee	
107 480	207 240	Plant filing fee	
108 690	208 345	Reissue filing fee	
114 150	214 75	Provisional filing fee	
SUBTOTAL (1)			(\$) 690.00

2. EXTRA CLAIM FEES

Total Claims	Extra Claims	Fee from below	Fee Paid
20	-20** = 0	X \$18.00 =	0.00
3	-3** = 0	X \$78.00 =	0.00
Multiple Dependent			

**or number of previously paid, if greater; For Reissues, see below

Large Entity Small Entity

Fee Code (\$)	Fee Code (\$)	Fee Description
103 18	203 9	Claims in excess of 20
102 78	202 39	Independent claims in excess of 3
104 270	204 135	Multiple Dependent claim
109 78	209 39	**Reissue independent claims over original patent
110 18	210 9	**Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) (\$) **0.00**

FEE CALCULATION (continued)

3. ADDITIONAL FEE

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description	Fee Paid
105 130	205 65	Surcharge - late filing fee or oath	
127 50	227 25	Surcharge - late provisional filing fee or cover sheet.	
139 130	139 130	Non-English specification	
147 2,520	147 2,520	For filing a request for reexamination	
112 920*	112 920*	Requesting publication of SIR prior to Examiner action	
113 1,840*	113 1,840*	Requesting publication of SIR after Examiner action	
115 110	215 55	Extension for response within first month	
116 380	216 190	Extension for response within second month	
117 870	217 435	Extension for response within third month	
118 1,360	218 680	Extension for response within fourth month	
128 1,850	228 925	Extension for response within fifth month	
119 300	219 150	Notice of Appeal	
120 300	220 150	Filing a brief in support of an appeal	
121 260	221 130	Request for oral hearing	
138 1,360	138 1,360	Petition to institute a public use proceeding	
140 110	240 55	Petition to revive - unavoidably	
141 1,210	241 605	Petition to revive - unintentionally	
142 1,210	242 605	Utility issue fee (or reissue)	
143 430	243 215	Design issue fee	
144 580	244 290	Plant issue fee	
122 130	122 130	Petitions to the Commissioner	
123 50	123 50	Petitions related to provisional applications	
126 240	126 240	Submission of Information Disclosure Stmt	
581 40	581 40	Recording each patent assignment per property (times number of properties)	
146 760	246 380	Filing a submission after final rejection (37 CFR 1.129(a))	
149 760	249 380	For each additional invention to be examined (37 CFR 1.129(b))	
Other fee (specify) _____			
Other fee (specify) _____			

* Reduced by Basic Filing Fee Paid

SUBTOTAL (3) (\$) _____

SUBMITTED BY

Typed or Printed Name **Farzad E. Amini, Reg. 42,261**

Signature _____

Date

09/29/00

Complete (if applicable)

Reg. Number

Deposit Account User ID

02-2666

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

UNITED STATES PATENT APPLICATION FOR

**RUNTIME PREDICTION FRAMEWORK FOR CPU INTENSIVE
APPLICATIONS**

Inventor(s):

**Ganapati Srinivasa
Hemanth Kumar
Donald Pearson
Mysore Sriram**

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP

12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025-1026
(310) 207-3800

006260" 98252960

RUNTIME PREDICTION FRAMEWORK FOR CPU INTENSIVE APPLICATIONS

Background

The invention relates generally to a method and system for providing the actual cost to the customer for utilization of a CPU farm, prior to that utilization.

In recent times, photo-realistic computer generated images have come to play a familiar role in commercial film and television productions. Production studios use sophisticated, compute-intensive software packages such as Maya™ from Alias Wavefront Corp., RenderMan™ from Pixar Corp., or Mental Ray™ from Mental Images, Inc. to create (or render) these images. Compute-intensive applications are characterized by their heavy use of a computer's central processing units. Depending on the complexity of a scene, a single frame can take up to several hours to render on conventional computer hardware. Since there are 1,800 frames in one minute of film (30 frames-per-second times 60), rendering time can quickly become impractical.

Faced with this dilemma, a production studio can either invest in additional computer hardware or outsource the rendering to a company (known as a "CPU farm") that has ample computer resources. Because the hardware investment can be significant and because there is a need to keep computers free for normal operations, smaller production studios are turning to CPU farms for rendering. A CPU farm can include a single computer system, but typically farms employ large numbers of fast processors interconnected through high-speed networks. The farm processors can be physically at one location or can be geographically dispersed. By dividing the rendering task among a number of processors, the farm can render a sequence of frames in a fraction of the time that it would take a small production studio.

Currently, CPU farms do not have the ability to quote the actual cost of performing a rendering job. Instead, farms provide an estimated cost. For example, the Alternative Perspective Render Farm (www.ap3d.com) extrapolates the time required for rendering based on the rendering time of one or more exemplary frames chosen by the customer. The calculation is based on the speed of the production studio's computer, the time it takes to render the exemplary frames, and the number of farm processors hired. Another farm, The Diner, Inc. (www.dinerinc.com), bases its cost estimation on the number of frames to be

rendered and the average rendering time per frame (also provided by the customer).

Typically, the actual cost is determined only after the farm has completed the rendering job. If the actual cost is greater than the estimated cost, the customer pays the difference. Without knowing the actual cost of rendering beforehand, customers may find it difficult to budget such expenditures.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is illustrated by way of example and not by way of limitation in the figures of the accompanying drawings in which like references indicate similar elements. It should be noted that references to "an" embodiment in this disclosure are not necessarily to the same embodiment, and they mean at least one.

Figure 1 depicts a block diagram of an exemplary computer system to implement certain embodiments of the present invention.

Figure 2 depicts a high-level view of an embodiment of a CPU farm.

Figure 3 depicts a software architecture for an embodiment of the method and system of the invention.

Figure 4 illustrates a flow of operations for an embodiment of the invention's method.

DETAILED DESCRIPTION

Referring to the figures, exemplary embodiments of the invention will now be described. The exemplary embodiments are provided to illustrate aspects of the invention and should not be construed as limiting the scope of the invention. The exemplary embodiments are primarily described with reference to diagrams. Depending upon the implementation, the corresponding apparatus element may be configured in hardware, software, firmware or combinations thereof.

Exemplary Computer Architecture

As discussed herein, a "system", such as the system for generating a scaled-down representation of a customer's data, or a system for calculating the turn-around time and actual cost to the customer to run a compute-intensive application on a CPU farm, may be an apparatus including hardware and/or

software for processing data. The system may include, but is not limited to, a computer (e.g., portable, laptop, desktop, server, mainframe, etc.), hard copy equipment (e.g., printer, plotter, fax machine, etc.), banking equipment (e.g., an automated teller machine), and the like.

5 A computer system **100** representing an exemplary workstation, host, or server in which features of the invention may be implemented will now be described with reference to **Figure 1**. The computer system **100** represents one possible computer system for implementing embodiments of the present invention, however other computer systems and variations of computer system
10 **100** are also possible. The computer system **100** has a bus or other communications channel **101** for communicating information, and a processor **102** coupled with bus **101** for processing information. The computer system **100** further includes a random access memory (RAM) or other dynamic storage device **104** (referred to as main memory), coupled to the bus **101** for storing information and instructions to be executed by the processor **102**. The main
15 memory **104** also may be used for storing temporary variables or other intermediate information during execution of instructions by the processor **102**. The computer system **100** also has a read only memory (ROM) and/or other static storage device **106** coupled to the bus **101** for storing static information and instructions for the processor **102**.
20

 A data storage device **107** such as magnetic disk, zip, or optical disc and its corresponding drive may also be coupled to the computer system **100** for storing information and instructions. The computer system **100** may also be coupled via the bus **101** to a display device **121**, such as a cathode ray tube (CRT)
25 or liquid crystal display (LCD), for displaying information to an end user. Typically, an alphanumeric input device such as a keyboard **122**, including alphanumeric and other keys, may be coupled to the bus **101** for communicating information and/or command selections to the processor **102**. Another type of user input device is a cursor control device **123**, such as a mouse, a trackball, or
30 cursor direction keys for communicating direction information and command selections to the processor **102** and for controlling cursor movement on the display **121**.

 A communication device **125** is also coupled to the bus **101**. Depending upon the particular implementation, the communication device **125** may include

a modem, a network interface card, or other well-known interface devices, such as those used for coupling to Ethernet, token ring, or other types of physical attachment for purposes of providing a communication link to support a local or wide area network, for example. In this manner, the computer system **100** may be coupled to a number of clients and/or servers via a conventional network infrastructure, such as a company's Intranet and/or the Internet, for example.

Exemplary CPU Farm

Figure 2 depicts a high-level view of an embodiment of the exemplary CPU farm. The farm has one or more individual computer systems **214**, also referred to here as 'processors for hire', that are connected by a network. The farm processors are hired by farm customers to execute compute-intensive applications. Scheduling the farm processors to perform customer work is handled by a farm control section **212**, which can include one or more systems dedicated to managing the farm. Interaction with the customer, including providing quotes and accepting work orders, is facilitated by one or more servers, also referred to in this embodiment as a Web head **208**.

The Web head **208** may consult with the farm control section **212** when providing cost and turn-around-time quotes to the customer. In addition, the Web head can include one or more systems and can host one or more Web servers. The customer's computer system (or client) can run a Web browser **200** which can be used to communicate with the farm Web head **208**. Communication with the farm can take place over the Internet, or any other wide area network **202**. Interposed between the Web head and the Internet are one or more firewalls **204, 206**.

In a typical transaction with the farm, a customer will use the Web browser **200** to contact the Web head **208** of the farm and submit a request to hire a number of farm processors **214** for running the customer's compute-intensive application. The farm may provide the customer a rough estimate on the cost to run the application. This estimate can be based on a number of factors, including the average frame processing time (as provided by the customer), the time it takes to render an exemplary frame (as provided by the customer), the number of frames, and the number of farm processors hired. Assuming the cost estimate is agreeable to the customer, the customer will then transfer any required input

data files to the farm. The farm control section **212** will then schedule individual processors **214** to run the customer's application with the provided input data. Once the run is finished, the customer can retrieve the results via the Web head **208**. If the actual time it took to run the customer's application was greater than the estimated time, the customer pays the difference.

Software Architecture

Figure 3 shows a diagram of a software architecture for and embodiment of the invention. The process of providing an accurate price quote to the customer is illustrated as a flow diagram in **Figure 4**. It begins with automatically analyzing the customer's input data on the customer's computer system using an Application-Specific Module (ASM) **300**, as indicated by block **400**. The ASM is tailored to a particular compute-intensive application. In this exemplary embodiment, the ASM is specific to a computer graphics rendering application, such as Maya™ or RenderMan™. However, the ASM **300** is not limited to computer graphics rendering applications. For example, the ASM could be tailored to an electronic logic simulator.

The ASM **300** scans the compute-intensive application's input data file(s) **302** and collects statistical information relevant to calculating the actual rendering time on the farm. This statistical information represents a scaled-down version of the application's input data, as referenced by block **402**. Two kinds of data are collected from the input data files: scene description and render options. Many of these parameters are discussed in general reference or widely available computer graphics texts such as *Computer Graphics Principles and Practice*, by James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes, Addison-Wesley Publishing Company, 1996 (ISBN 0201848406). Others may be specific to certain graphics software packages, in which case they may be discussed in the documentation for that package.

The scene description is a three dimensional model of the graphic to be rendered. It includes the geometry that describes the objects in the scene as well as any lights that illuminate the scene, the number of triangles needed to represent the scene, textures, shading method, and a camera used to capture the scene in two dimensions. Lights may be provided in the three dimensional model to illuminate objects. Lights may have different colors, intensities, and

directions. Objects may have reflective or absorptive properties for light. Textures and shading methods are used to create effects on the surfaces of objects. Shading methods can include flat, Gouraud, or Phong, for example. These elements are used to determine the base complexity (or base render time) of a render job.

The render options are the parameters that render engines use to improve the quality and speed of a render job. Some examples include resolution, ray-tracing, anti-aliasing, and motion-blur. Ray tracing simulates the path of a light ray as it is reflected and absorbed by objects in the image. Anti-aliasing refers to the process of reducing jagged distortions in curves and diagonal lines. Motion-blur is an effect that simulates a camera shutter remaining open for an extended period of time, giving objects the appearance of movement. As more options are selected, the scene's complexity, and hence render time, increases.

The scaled-down data is then transmitted over a computer network (e.g., the Internet) to the CPU farm. On the farm, a heuristic modeler 306 (also referenced as block 404 in Fig. 4) consumes the scaled-down data and calculates a computing time for the class of applications (e.g., rendering or logic simulation).

The computing time, also referred to as the parameterized cost 308 (also referenced as block 406 in Fig. 4), is a product of the base complexity, motion-blur/anti-aliasing heuristic, and the ray tracing heuristic. The base complexity is calculated using the number of triangles in the scene and the frame resolution (pixels). The motion-blur/anti-aliasing heuristic takes into account the additional time required to render frames using anti-aliasing and/or motion-blur algorithms. The ray tracing heuristic is a multiplier based on a linear relationship between the frame resolution and the number of light sources.

The run-time correlator 310 takes the computing time calculated by the heuristic modeler 306 and predicts, as referenced by block 408, the actual cost and turn-around time by taking into account the set of available processors, the processor speeds, and whether or not the customer needs the results as soon as possible. The actual cost and turn-around time 410 is then transmitted over a computer network 312 to the customer system as part of a quote 314. If the compute-intensive application is for graphics rendering, the cost could be in terms of cost per frame. Likewise, if the compute-intensive application is for logic simulation, the cost could be in terms of cost per gate.

In the foregoing specification, the invention has been described with reference to specific exemplary embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

5

CLAIMS

We claim:

1 1. A method comprising:
2 creating a scaled-down representation of input to a compute-intensive
3 application;
4 calculating a computing requirement based on the scaled-down
5 representation;
6 calculating a turn-around time and an actual cost to a customer to run the
7 compute-intensive application with the input, on one or more processors, based
8 on the calculated computing requirement; and
9 sending the turn-around time and the actual cost to the customer's client
10 software.

1 2. The method of claim 1 wherein the compute-intensive application
2 is to perform computer graphics rendering.

1 3. The method of claim 1 wherein the compute-intensive application
2 is to perform logic simulation.

1 4. The method of claim 1 wherein the scaled-down representation of
2 the application input is generic to a class of applications.

1 5. The method of claim 1 wherein the scaled-down representation of
2 the application input includes the geometry, lights, number of triangles, textures,
3 shading method, camera, ray-tracing, anti-aliasing, and motion-blur of an
4 underlying scene.

1 6. The method of claim 1 further wherein the turn-around time and
2 actual cost are transmitted over an internet to the customer's client software.

1 7. The method of claim 1 wherein the cost is in terms of input units.

1 8. The method of claim 7 wherein the input units are logic gates.

1 9. The method of claim 7 wherein the input units are image frames.

1 10. A system comprising:

2 an application-specific module to model input data;

3 a heuristic modeler module coupled to the output of the application-
4 specific module, to calculate a computing requirement; and

5 a run-time calculator module coupled to the output of the heuristic
6 modeler module, to compute a turn-around time and an actual cost to run the
7 application on one or more processors.

1 11. The system of claim 10 wherein the modules are to communicate
2 with each other over an internet.

1 12. The system of claim 10 wherein the application-specific module is
2 to generate a scaled-down representation of the data to include the geometry,
3 lights, number of triangles, textures, shading method, camera, ray-tracing, anti-
4 aliasing, and motion-blur of an underlying scene.

1 13. An article of manufacture comprising:

2 a machine readable medium containing instructions which, when
3 executed by a processor, cause a machine to perform operations comprising:

4 calculating a computing requirement based on a scaled-down
5 representation of input to a compute-intensive application, the representation
6 having been created at a customer's machine;

7 calculating a turn-around time and an actual cost to the customer to run
8 the compute-intensive application with the input, on one or more processors,
9 based on the calculated computing requirement; and

10 providing the turn-around time and the actual cost to the customer's client
11 software.

1 14. The article of manufacture of claim 13 wherein the medium
2 includes further instructions to create the scaled-down representation of the
3 application input as being generic to a class of applications.

1 15. The article of manufacture of claim 13 wherein the medium
2 includes further instructions to create the scaled-down representation of the
3 application input as having the geometry, lights, number of triangles, textures,
4 shading method, camera, ray-tracing, anti-aliasing, and motion-blur of an
5 underlying scene.

1 16. The article of manufacture of claim 13 wherein the medium
2 includes further instructions to enable the scaled-down representation of the
3 input to be received over an internet from the client software.

1 17. The article of manufacture of claim 13 wherein the medium
2 includes further instructions to enable the turn-around time and actual cost to be
3 transmitted over the internet to the customer's client software.

1 18. The article of manufacture of claim 13 wherein the medium
2 includes further instructions to calculate the cost in terms of input units.

1 19. The article of manufacture of claim 18 wherein the medium
2 includes further instructions to calculate the cost in terms of input units being
3 logic gates.

1 20. The article of manufacture of claim 18 wherein the medium
2 includes further instructions to calculate the cost in terms of input units being
3 image frames.

ABSTRACT

A scaled-down representation of input to a compute-intensive application is created. A computing requirement based on the scaled-down representation is calculated. A turn-around time and an actual cost to a customer to run the compute-intensive application with the input, on one or more processors, based on the calculated computing requirement, is calculated and then sent to the customer.

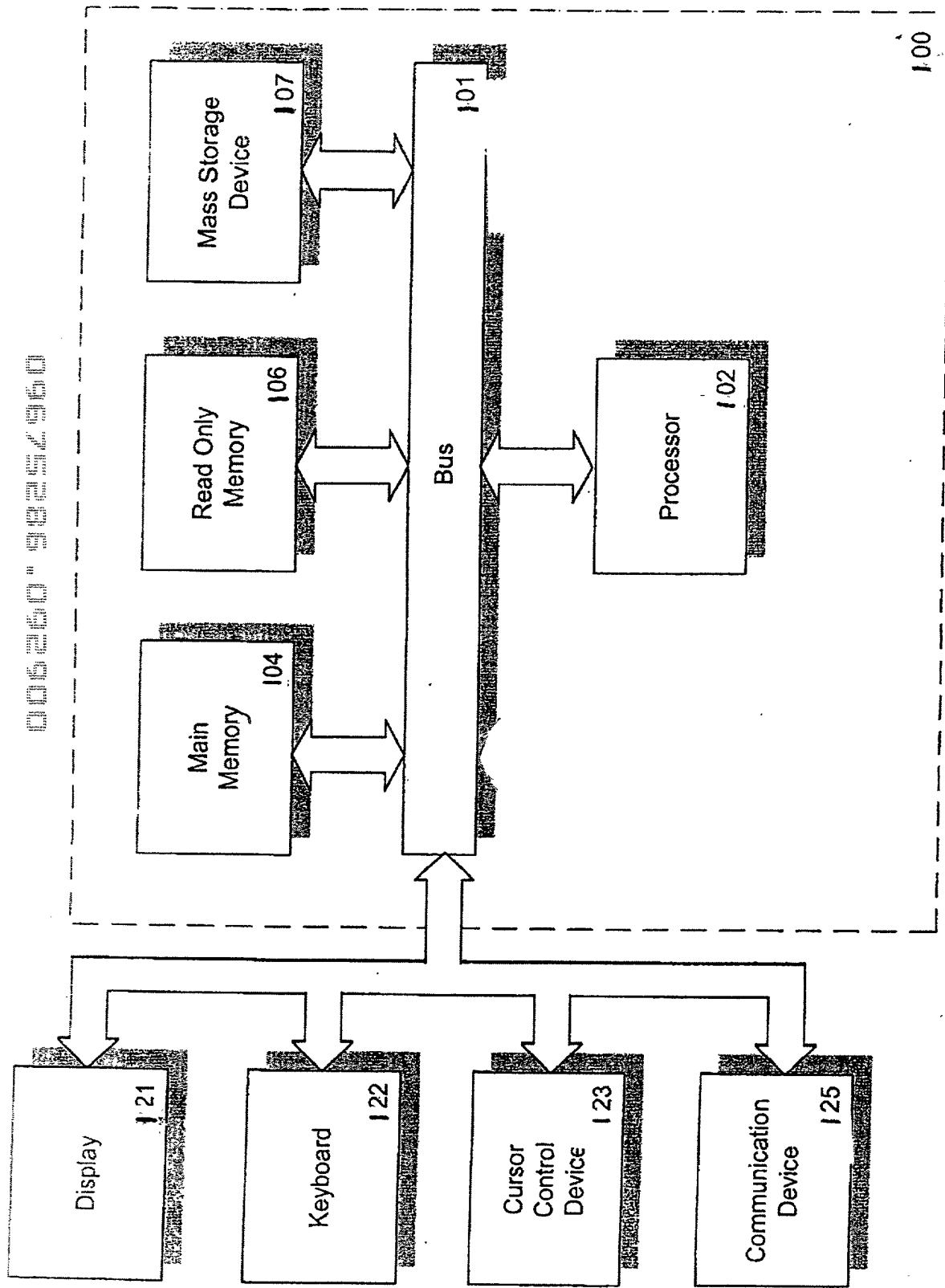


Figure 1

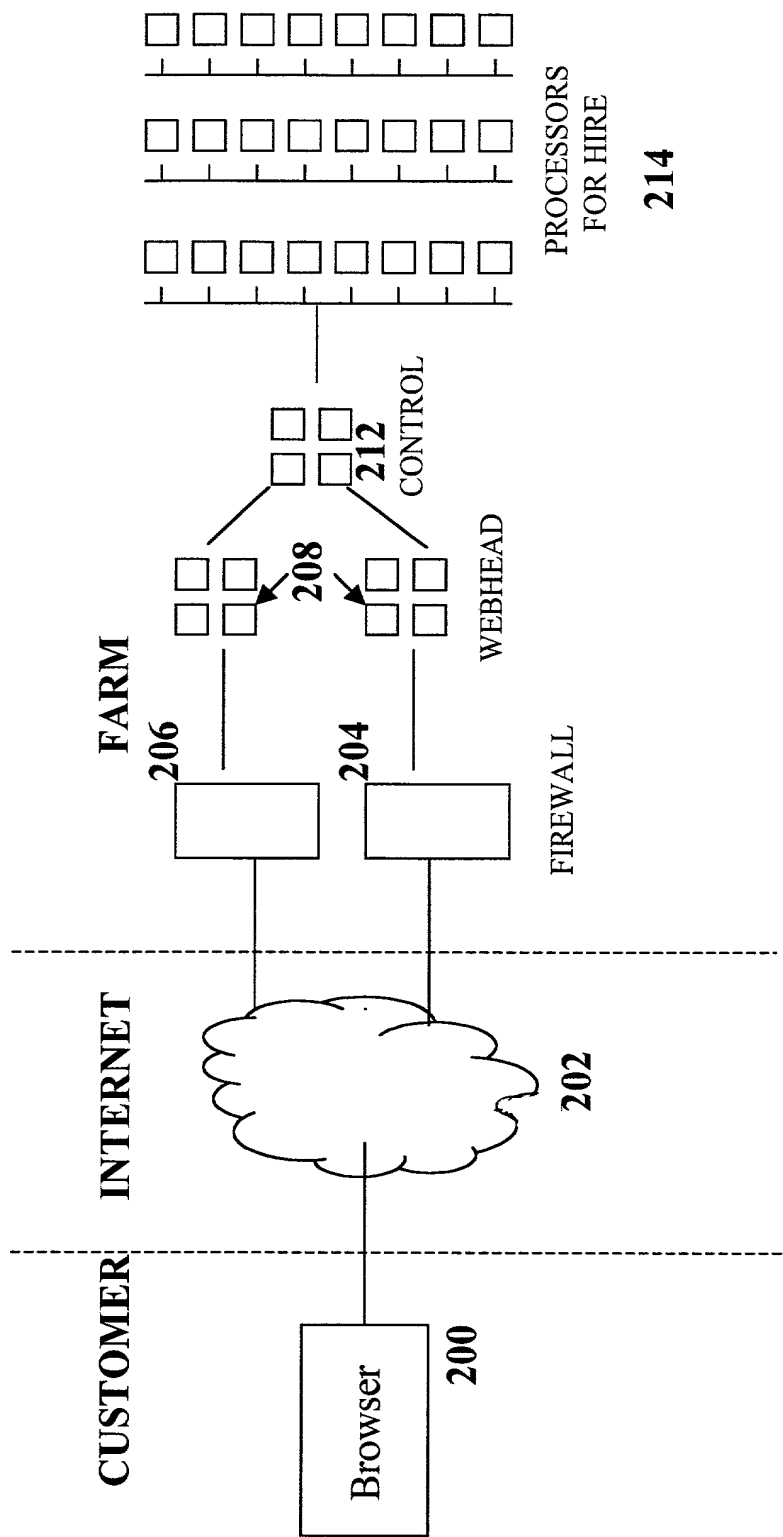


Fig: 2

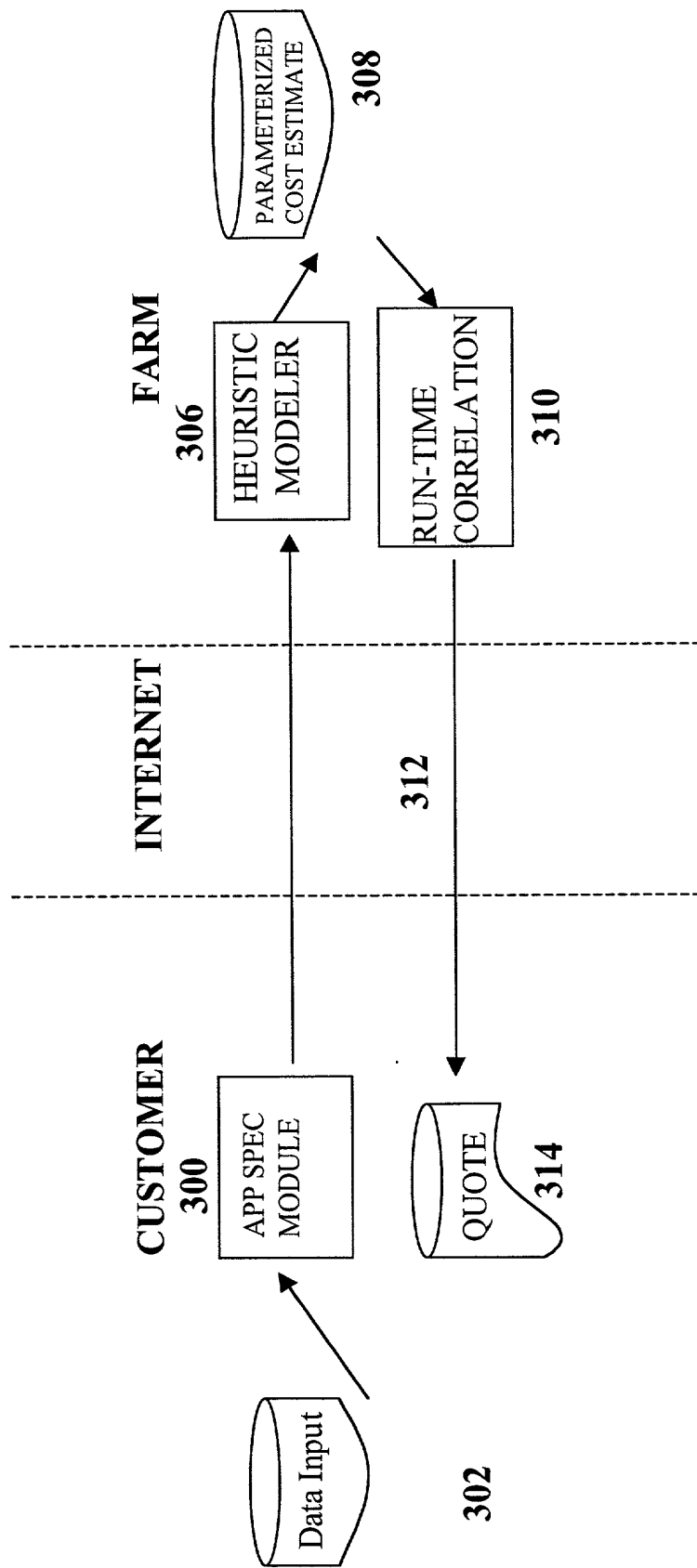


Fig: 3

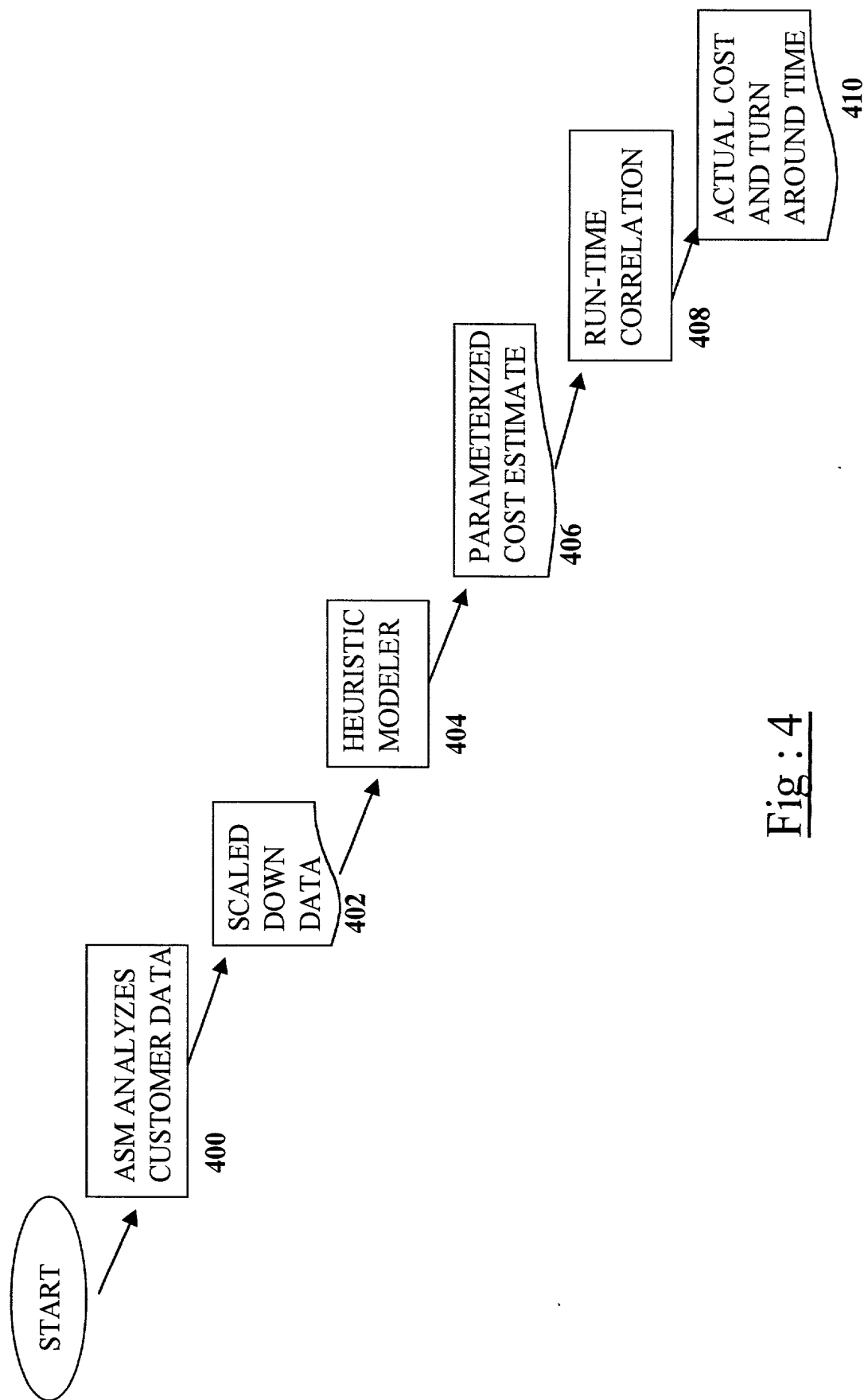


Fig : 4